

**PROJECT TITLE:**

# **VISUALIZING AND PREDICTING HEART DISEASE WITH AN INTERACTIVE DASHBOARD**

**TEAM ID:** PNT2022TMID15265

**TEAM MEMBERS:**

KOUSHIGAN L (TEAM LEADER)

KIRAN S

KOTHURI SAI SUMANTH

LOGESHWARAN Y S

# 1.INTRODUCTION

## 1.1 PROJECT OVERVIEW

Heart Attack Prediction System is an initiative done to notify people about their heart conditions.

During a heart attack, the blood supply that normally nourishes the heart with oxygen is cut off and the heart muscle begins to die. Heart attacks also called myocardial infarctions are very common in the United States. In fact, it's estimated that one happens every 40 seconds Trusted Source.

Some people who are having a heart attack have warning signs, while others show no signs. Some symptoms that many people report are:

- chest pain
- upper body pain
- sweating
- nausea
- fatigue
- trouble breathing

A heart attack is a serious medical emergency. Seek immediate medical attention if you or someone you know is experiencing symptoms that could signal a heart attack.

Causes for heart attack:

There are a few cardiac conditions that can cause heart attacks. One of the most common causes is plaque buildup in the arteries (atherosclerosis) that prevents blood from getting to the heart muscle.

Heart attacks can also be caused by blood clots or a torn blood vessel. Less commonly, a heart attack is caused by a blood vessel spasm.

Predicting and diagnosing heart disease is the biggest challenge in the medical industry and it is based on factors like physical examination, symptoms and signs of the patient. Factors which influence heart diseases are cholesterol level of the body, smoking habit, and obesity, family history of diseases, blood pressure and working environment. Machine learning algorithms play a vital and accurate role in predicting heart disease. The advancement of technologies allows machine language to pair with big data tools to handle unstructured and exponentially growing data. The leading cause of death in the developed world is heart disease. Therefore, there needs to be work done to help prevent the risks of having a heart attack or stroke..

## 1.2 PURPOSE

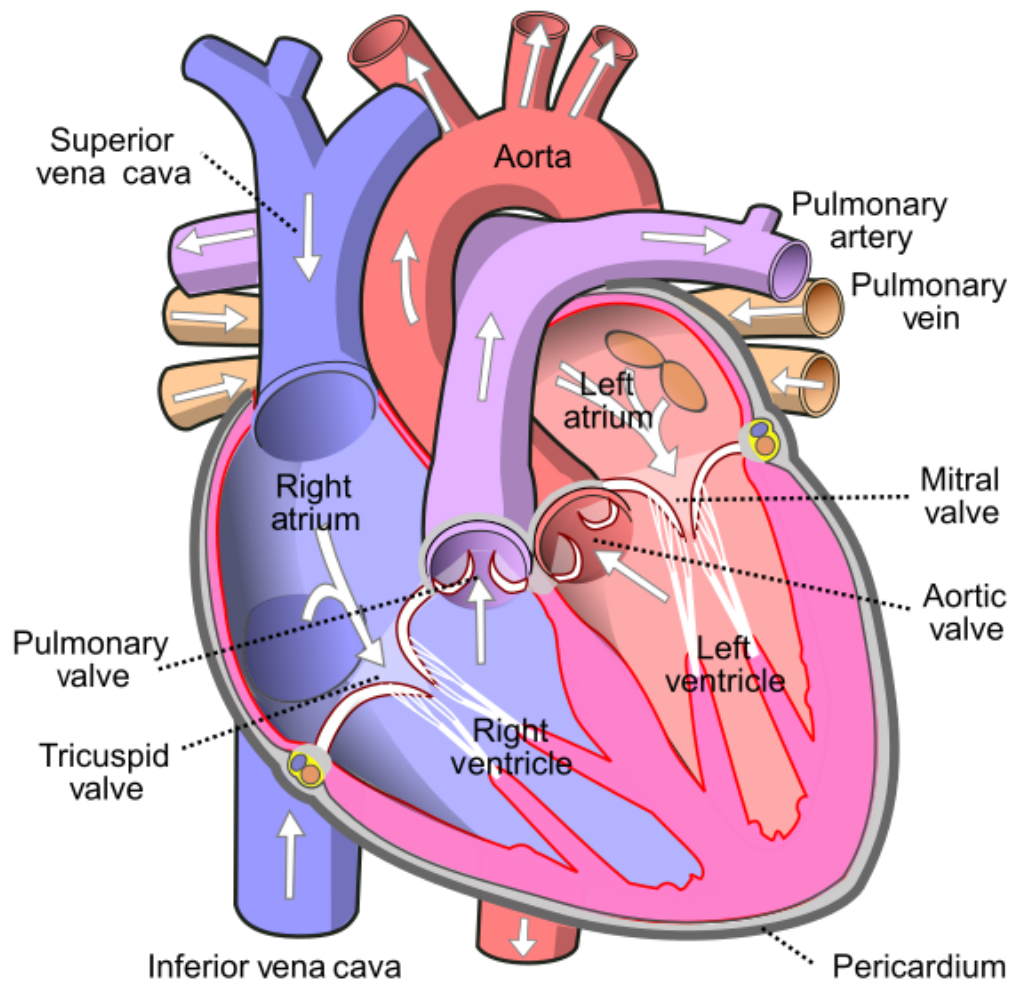
The heart is important because it pumps blood around your body, delivering oxygen and nutrients to your cells and removing waste products. The heart has four chambers: two upper chambers called the right and left atria (singular: atrium) and two lower chambers called the right and left ventricles.

There are valves between the atria and ventricles that make sure blood flows in one direction through your heart.

The four main functions of the heart are:

- Pumping oxygenated blood to other body parts.
- Pumping hormones and other vital substances to different parts of the body.
- Receiving deoxygenated blood and carrying metabolic waste products from the body and pumping it to the lungs for oxygenation.
- Maintaining blood pressure.

These much importance provided organ part of our body need to be precautionarily cared without letting it to affect any diseases or attacks. Using this heart attack prediction system one can find the probability of getting the heart attack also the factors which predominantly causes attack for the user. By these outputs precautionary methods can be made by the user and save his/her life.



Heart disease can be prevented by following a healthy living. Below mentioned are few heart attack precaution strategies and tips to prevent a heart attack. A healthy lifestyle change reduces the risk of heart diseases. Let's see some easy lifestyle changes to prevent heart attacks:

- Quit smoking
- Keep check and control the cholesterol levels
- Control blood pressure levels

- Exercise and be physically fit
- Loose the extra weights
- Choose a diet for a healthy heart
- Get proper sleep
- Manage stress
- Control the blood sugar level

The leading cause of death in the developed world is heart disease. Therefore, there needs to be work done to help prevent the risks of having a heart attack or stroke. Using the dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features available.

## **2.LITERATURE SURVEY**

### **2.1 EXISTING PROBLEM**

A major challenge faced by health care organizations, such as hospitals and medical centers, is the provision of quality services at affordable costs.<sup>1</sup> The quality service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the database.<sup>2</sup> The proposed system can determine an exact hidden knowledge, ie, patterns and relationships associated with heart disease from a historical heart disease database. It can also answer the complex queries for diagnosing heart disease; therefore, it can be helpful to health care practitioners to make intelligent clinical decisions. Results showed that the proposed system has its unique potency in realizing the objectives of the defined mining goals.

### **2.2 PROBLEM STATEMENT DEFINITION**

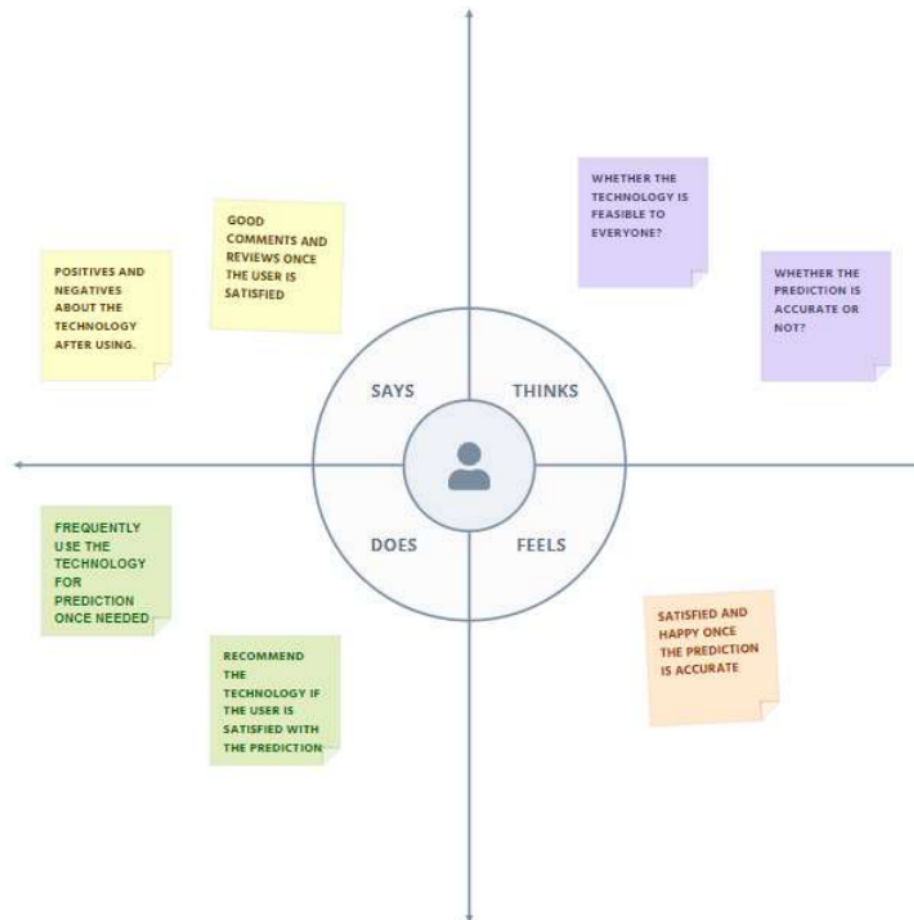
Heart acts a major role in the corporeal organisms. The diseases of the heart want more perfection and exactness for diagnosis and analyses. Heart disease is a

dangerous disease. This disease occurs due to various problems such as overpressure, blood sugar, high blood pressure, Cholesterol, etc. in the human body By using Python and machine learning, this paper is analyzed and predicted heart disease. We can predict this disease by using various attributes in the data set. We have collected a data set consisting of 13 elements and 383 individual values to analyze the patient's performance. The main aim of the paper is to get better accuracy to detect heart disease using the ML algorithm.

<b>Problem Statement(PS)</b>	<b>I am (Customer)</b>	<b>I'm trying to</b>	<b>But</b>	<b>Because</b>	<b>Which makes me feel</b>
PS-1	Cust1	Check if I have any heart disease	I have Diabetes	Age factor	Stressed
PS-2	Cust2	Check if I have any heart disease	I am completely healthy	Healthy Diet	Relieved

### 3. IDEATION & PROPOSED SOLUTION

#### 3.1 EMPATHY MAP CANVAS

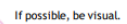


A psychologist with a lifestyle issue during a consultation. patient

#### 3.2 IDEATION AND BRAIN STORMING

In order to make the dataset we are working with easier to understand, we will first take a look at data wrangling. A timely diagnosis of heart disease (HD), one of the most prevalent diseases today, is essential for many healthcare professionals to protect their patients from the condition and save lives. To accurately classify and/or predict HD cases with a few variables, a comparison analysis of various classifiers can be done for the classification of the heart disease dataset.

Accurate decision-making and ideal therapy are needed to address cardiac risk. Five machine learning models were utilized in a Canadian study to examine 1-month mortality among hospitalized patients with congestive heart failure. Several tests, including auscultation, blood pressure, cholesterol, ECG, and blood sugar, are carried out prior to the diagnosis of a condition. These tests assist in identifying the patient's medication requirements. In this work, the predictive accuracy of various machine learning methods is investigated to calculate cardiovascular risk. The performance comparison of the most recent REP Tree and Random Tree machine learning algorithms in terms of cardiovascular disease prediction is innovative.





2

## Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

**TIP**  
You can select a sticky note and hit the pencil (switch to sketch) icon to start drawing!



3

## Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

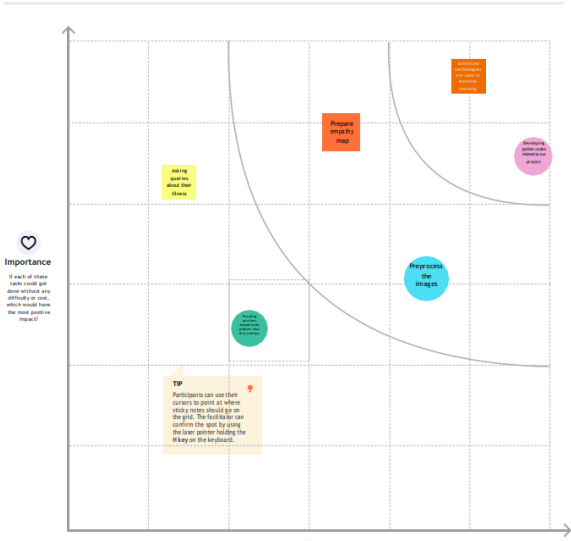
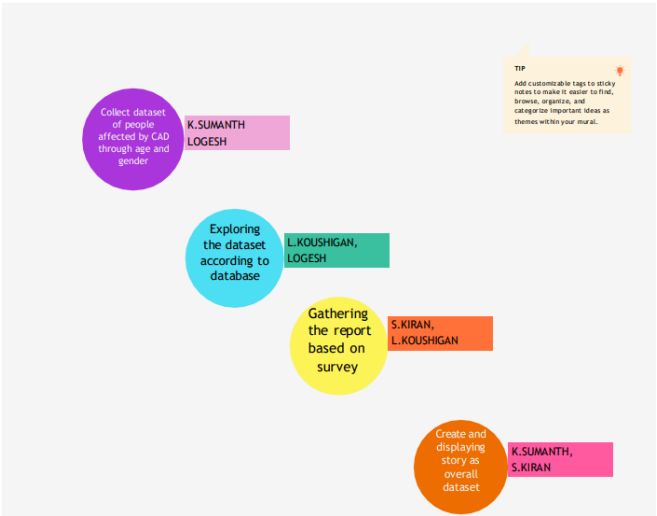
20 minutes

4

## Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes





## After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

---

### Quick add-ons



#### Share the mural

Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.



#### Export the mural

Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

---

### Keep moving forward



#### Strategy blueprint

Define the components of a new idea or strategy.

[Open the template](#) →



#### Customer experience journey map

Understand customer needs, motivations, and obstacles for an experience.

[Open the template](#) →



#### Strengths, weaknesses, opportunities & threats

Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.

[Open the template](#) →



[Share template feedback](#)

## 3.3 PROPOSED SOLUTION

S.NO	Parameter	Description
1.	Problem Statement(Problem to be solved)	1. To develop an interactive dashboard to predict the heartdisease accurately with few tests and attributes the presence of heart disease.
2.	Idea / Solution description	1. Analyzing data and identifying the heartdisease using Cognos analytics tools.
3.	Novelty / Uniqueness	1. Hoping to achieve maximum accuracy to provide prior treatment to the patients and reduce the fatality rate.
4.	Social Impact/ Customer Satisfaction	<p>1. Saving lives, User friendly interactive dashboard.</p> <p>2. Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions and experiences.</p>
5.	Business Model (Revenue Model)	<p>1. Data security.</p> <p>2. Easy to use.</p> <p>3. Constant updates according to necessity.</p>
6.	Scalability of the Solution	<p>1. Can be used in any platform (Windows, mac, etc.,)</p> <p>2. Adding new feature doesn't affect the performance of the system.</p> <p>3. Scalable dataset.</p>

## ALGORITHMS USED

## Logistic Regression:

Logistic regression is used in various fields, including machine learning, most medical fields, and social sciences. For example, the Trauma and Injury Severity Score ), which is widely used to predict mortality in injured patients, was originally developed by using logistic regression. Many other medical scales used to assess severity of a patient have been developed using logistic regression. Logistic regression may be used to predict the risk of developing a given disease based on observed characteristics of the patient (age, sex, bmi, results of various blood tests, etc.). Another example might be to predict whether a Nepalese voter will vote Nepali Congress or Communist Party of Nepal or Any Other Party, based on age, income, sex, race, state of residence, votes in previous elections, etc. The technique can also be used in engineering, especially for predicting the probability of failure of a given process, system or product. It is also used in marketing applications such as prediction of a customer's propensity to purchase a product or halt a subscription, etc. In economics it can be used to predict the likelihood of a person ending up in the labor force, and a business application would be to predict the likelihood of a homeowner defaulting on a mortgage. Conditional random fields, an extension of logistic regression to sequential data, are used in natural language processing.

```
# Logistic Regression

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn import metrics
from sklearn.metrics import roc_curve

logit = LogisticRegression()
logit.fit(X_train, y_train)

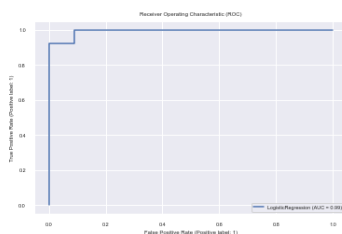
predicted_logit = logit.predict(X_test)

LogisticRegressionScore = accuracy_score(predicted_logit, y_test)

plt.figure()
metrics.plot_roc_curve(logit, X_test, y_test)
plt.title("Receiver Operating Characteristic (ROC)")
plt.show()

print("Logistic Regression score: ", LogisticRegressionScore)
```

c:\users\91735\appdata\local\programs\python\python39\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning:  
Function plot\_roc\_curve is deprecated; Function :func:'plot\_roc\_curve' is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods:  
:meth:'sklearn.metrics.RocCurveDisplay.from\_predictions' or :meth:'sklearn.metrics.RocCurveDisplay.from\_estimator'.



Logistic Regression score: 0.9166666666666666

## Gaussian Naive Bayes:

This algorithm offers plenty of advantages to its users. That's why it has a lot of applications in various sectors too.

As this algorithm is fast and efficient, you can use it to make real-time predictions.

This algorithm is popular for multi-class predictions. You can find the probability of multiple target classes easily by using this algorithm.

Email services (like Gmail) use this algorithm to figure out whether an email is a spam or not. This algorithm is excellent for spam filtering.

Its assumption of feature independence, and its effectiveness in solving multi-class problems, makes it perfect for performing Sentiment Analysis. Sentiment Analysis refers to the identification of positive or negative sentiments of a target group (customers, audience, etc.)

Collaborative Filtering and the Naive Bayes algorithm work together to build recommendation systems. These systems use data mining and machine learning to predict if the user would like a particular resource or not.

```
# Gaussian Naive Bayes

from sklearn.naive_bayes import GaussianNB

gauss = GaussianNB()
gauss.fit(X_train, y_train)

gauss_pred = gauss.predict(X_test)

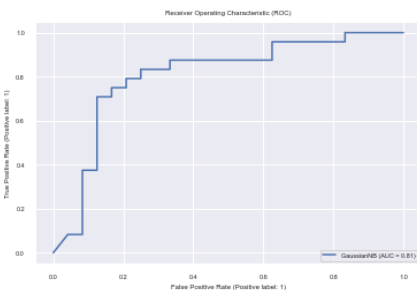
gauss_score = accuracy_score(gauss_pred, y_test)

plt.figure()
metrics.plot_roc_curve(gauss, X_test, y_test)
plt.title("Receiver Operating Characteristic (ROC)")
plt.show()

print("Gaussian Naive Bayes score: ", gauss_score)
```

c:\users\91735\appdata\local\programs\python\python39\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning:

Function plot\_roc\_curve is deprecated; Function :func:`plot\_roc\_curve` is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: :meth:`sklearn.metrics.RocCurveDisplay.from\_predictions` or :meth:`sklearn.metrics.RocCurveDisplay.from\_estimator`.



Gaussian Naive Bayes score: 0.7708333333333334

## KNeighborsClassifier:

The KNN algorithm can compete with the most accurate models because it makes highly accurate predictions. Therefore, you can use the KNN algorithm for applications that require high accuracy but that do not require a human-readable model.

The quality of the predictions depends on the distance measure. Therefore, the KNN algorithm is suitable for applications for which sufficient domain knowledge is available. This knowledge supports the selection of an appropriate measure.

The KNN algorithm is a type of lazy learning, where the computation for the generation of the predictions is deferred until classification. Although this method increases the costs of computation compared to other algorithms, KNN is still the better choice for applications where predictions are not requested frequently but where accuracy is important.

```
# KNeighborsClassifier
from sklearn.neighbors import KNeighborsClassifier

KNC = KNeighborsClassifier(n_neighbors=2)
KNC.fit(X_train, y_train)

KNC_pred = KNC.predict(X_test)

KNC_accuracy = metrics.accuracy_score(y_test, KNC_pred)

print("KNeighbourClassifier score: ", KNC_accuracy)

KNeighbourClassifier score:  0.7916666666666666
```

## Random Forest Classifier:

Random forests is a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

Random forests has a variety of applications, such as recommendation engines, image classification and feature selection. It can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases. It lies at the base of the Boruta algorithm, which selects important features in a dataset.

```
# Random Forest Classifier

from sklearn.ensemble import RandomForestClassifier

rnd_clf = RandomForestClassifier(n_estimators=500, max_leaf_nodes=16, n_jobs=-1)
rnd_clf.fit(X_train, y_train)

rnd_clf_pred = rnd_clf.predict(X_test)

rnd_clf_accuracy = metrics.accuracy_score(y_test, rnd_clf_pred)
print("RandomForest score: ", rnd_clf_accuracy)
```

RandomForest score: 0.8333333333333334

## Bagging Decision Tree:

Bagging, also known as Bootstrap aggregating, is an ensemble learning technique that helps to improve the performance and accuracy of machine learning algorithms. It is used to deal with bias-variance trade-offs and reduces the variance of a prediction model. Bagging avoids overfitting of data and is used for both regression and classification models, specifically for decision tree algorithms.

```
# Bagging Decision Tree

from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier

bag_clf = BaggingClassifier(
    DecisionTreeClassifier(), n_estimators = 500, oob_score=True,
    max_samples=100, bootstrap = True, n_jobs=-1)

bag_clf.fit(X_train, y_train)

bag_clf_oob = bag_clf.oob_score_

bag_clf_pred = bag_clf.predict(X_test)
bag_clf_accuracy = metrics.accuracy_score(y_test, bag_clf_pred)
print("Bagging Decision Tree score: ", bag_clf_accuracy, "Out of the bag: ", bag_clf_oob)
```

Bagging Decision Tree score: 0.8541666666666666 Out of the bag: 0.7760416666666666

## Hard Voting Classifier:

Hard voting classifier classifies input data based on the **mode** of all the predictions made by different classifiers. The majority voting is considered differently when weights associated with the different classifiers are equal or otherwise.

```
# Hard Voting Classifier

from sklearn.ensemble import VotingClassifier

voting_clf = VotingClassifier(
    estimators=[('gauss', gauss), ('logit', logit), ('KN', KNC), ('bag', bag_clf), ('rnd_clf', rnd_clf)],
    voting = 'soft'
)

voting_clf.fit(X_train, y_train)

voting_clf_pred = bag_clf.predict(X_test)
voting_clf_accuracy = metrics.accuracy_score(y_test, voting_clf_pred)
print("Voting Classifier score: ", voting_clf_accuracy)
```

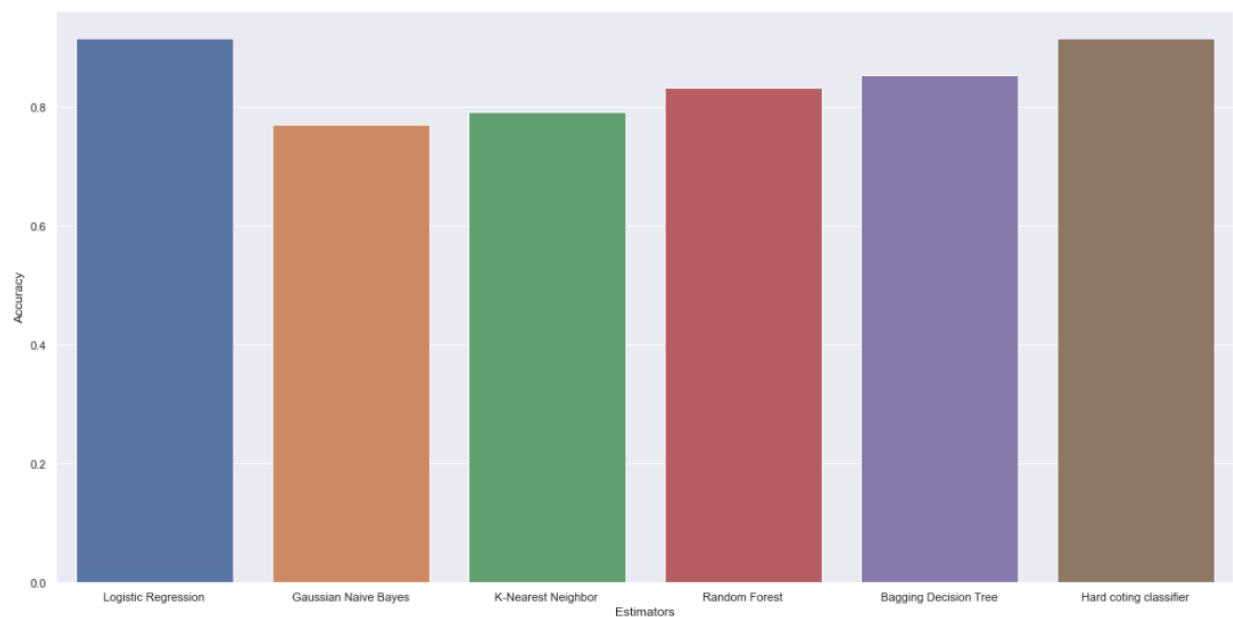
Voting Classifier score: 0.9166666666666666

## Model Accuracy

```
data = {'Estimators': ['Logistic Regression',  
                      'Gaussian Naive Bayes',  
                      'K-Nearest Neighbor',  
                      'Random Forest',  
                      'Bagging Decision Tree',  
                      'Hard coting classifier'],  
        "Accuracy": [LogisticRegressionScore,  
                      gauss_score,  
                      KNC_accuracy,  
                      rnd_clf_accuracy,  
                      bag_clf_accuracy,  
                      voting_clf_accuracy]  
      }  
  
data = pd.DataFrame(data)  
data.sort_values('Accuracy', ascending=False)
```

	Estimators	Accuracy
0	Logistic Regression	0.916667
5	Hard coting classifier	0.916667
4	Bagging Decision Tree	0.854167
3	Random Forest	0.833333
2	K-Nearest Neighbor	0.791667
1	Gaussian Naive Bayes	0.770833

By comparing the algorithms Logistic Regression high efficiency for our application, Hence it is used for our prediction process.





## **3.4 PROBLEM SOLUTION FIT**

### **3.4.1. Customer Segment**

All adults. Especially people who are older than 40 years and those who are on the verge of getting heart disease due to various factors such as age, obesity, diabetes, stress, etc.

### **3.4.2. Jobs-To-Be-Done / Problems J&P**

To predict and identify the heart disease patient. It is a very useful strategy that was used to control how the model can be utilized to increase the accuracy of the prediction of Heart Attack in each.

### **3.4.3. Triggers TR**

The generation currently living now leads an extremely unhealthy lifestyle. People worry about the sharp rise in mortality from heart-related illnesses. They, therefore, desire to adopt a better lifestyle.

### **3.4.4. Emotions: Before / After EM**

People frequently worry that their health will decline. They suffer unneeded tension and emotional breakdowns as a result of this. Our prediction system would enable them to keep track of their health independently and assist them in overcoming their erroneous concerns.

### **3.4.5. Available Solutions**

EDA: Exploratory data analysis is the key step for getting meaningful results.

Pros: Improve understanding of variables by extracting averages, mean, minimum, and maximum values, etc. Discover errors, outliers, and missing values in the data. Identify patterns by visualizing data in graphs such as box plots, scatter plots, and histograms.

Cons: Exploratory research comes with disadvantages that include offering inconclusive results, lack of standardized analysis, a small sample population, and outdated

information that can adversely affect the authenticity of the information.

#### **3.4.6. Customer Constraints**

The patient's medical status must be continuously monitored. Unpredictability could lead to inaccurate results. The patient must be genuine about the periodic readings they record. The process could consume the internet and could be slightly expensive.

#### **3.4.7. Behaviour**

To solve their problem, a suitable application must be available. To effectively diagnose the situation for their present health status, appropriate information such as age, weight, current symptoms, and cholesterol should be provided.

#### **3.4.8. Channels of Behaviour**

##### **3.4.8.1 ONLINE**

- Data Collected from offline devices is used in this application in order to visualize and predict heart diseases

##### **3.4.8.2 OFFLINE**

- ECG • Blood Sugar Level • Blood Pressure • Cholesterol

#### **3.4.9. Problem Root Cause**

The risk of heart disease is influenced by a number of variables, including smoking, body cholesterol, family history of the disease, obesity, high blood pressure, and inactivity.

## **4.REQUIREMENT ANALYSIS**

**Functional Requirements:**

Following are the functional requirements of the proposed solution.

<b>FR No.</b>	<b>Functional Requirement(Epic)</b>	<b>Sub Requirement (Story / Sub-Task)</b>
FR-1	User Registration	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

### **Non-functional Requirements:**

Following are the non-functional requirements of the proposed solution.

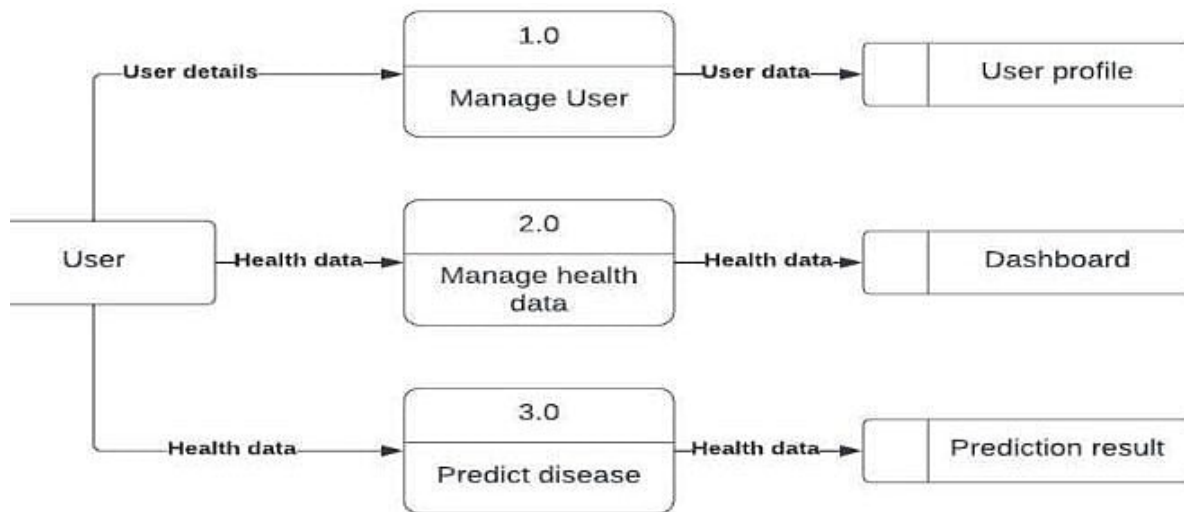
<b>NFR No.</b>	<b>Non-Functional Requirement</b>	<b>Description</b>
NFR-1	<b>Usability</b>	The application will have a simple and user-friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	<b>Security</b>	For security of the application the technique known as database replication should be used so that all the important data should be kept safe. In case of crash, the system should be able to backup and recover the data
NFR-3	<b>Reliability</b>	The application has to be consistent at every scenario and has to work without failure in any environment
NFR-4	<b>Performance</b>	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which

		depends on the efficiency of implemented algorithm
NFR-5	<b>Availability</b>	The application has to be available 24 x 7 for users without any interruption
NFR-6	<b>Scalability</b>	The application can withstand the increase in the no. of users and has to be able to develop High versions

## 5. PROJECT DESIGN

### 5.1 DATA FLOW DIAGRAMS

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



### 5.2 SOLUTION ARCHITECTURE

The solution architecture is as followed:

1. User registration
2. Questionnaires
3. Checking the probability
4. Generating the report

#### User Registration

Logging in, (or logging on or signing in or signing on), is the process by which an individual gains access to a computer system by identifying and authenticating themselves. The user credentials are typically some form of “username” and a matching “password”, and these credentials themselves are sometimes referred to as a login.

#### Questionnaires

Here the user feed the values in the application form, he/she fills up each and every details in the form. All these details gets saved in the server and details and from that we can extract the features of the disease. The entered details are matched with the datasets which are saved in the database.

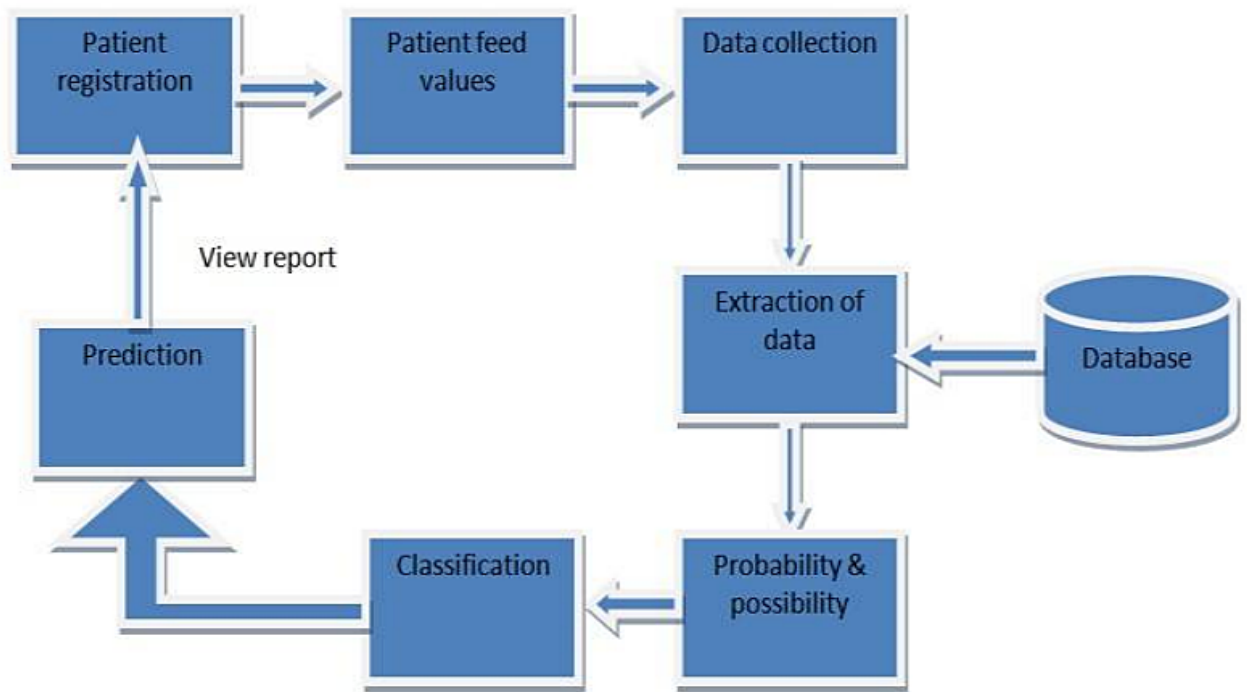
#### Checking for disease symptoms

After matching the details with the datasets it checks for the disease symptoms. One feature may match with different disease. So, it's necessary to check each and every matched details in order to predict the correct disease.

#### Generate Report

A report is being generated based on the matched symptoms. It predicts the disease and send it to user mobile application, and finally add some tips/suggestions to the user like nearby hospital details and it notifies patient by sending a message alert to patient mobile number.

### 5.3 TECHNICAL ARCHITECTURE



Initially the patient registers by providing certain parameters. That registered data is collected in a database by using machine learning techniques like data collection techniques and when he went to check about his health condition the collected values or data that has been stored in the database is been extracted by using some feature extraction techniques. When data is extracted, it under goes certain processes and therefore finally a disease is predicted and a report is generated. This is the overview of the heart disease prediction system using machine learning techniques.

**Table-1 : Components & Technologies:**

<b>S. No</b>	<b>Component</b>	<b>Description</b>	<b>Technology</b>
1.	User Interface	How user interacts with application e.g.Web UI, Mobile App, etc.	HTML, CSS, Python,Flask, Angular etc.
2.	Application Logic-1	Logic for a process in the application	Python
3.	Application Logic-2	Logic for a processin the application	IBM CognosAnalytics
4.	Application Logic-3	Logic for a processin the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	MongoDB, etc.
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Pak etc.
7.	File Storage	File storage requirements	Use Professional Records Storage, IBM Block Storage or Other StorageServices.
8.	External API	Purpose of External API used in theapplication	IBM SPSS,etc.
9.	Infrastructure (Server / Cloud)	Application Deployment on Local System/Cloud Local Server Configuration: Cloud ServerConfiguration :	Personal Server, IBM Cloud Serveretc.

**Table-2: Application Characteristics:**

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Open-source frameworks used	Technology of Opensource framework – Flask inPython and



			Angular.
2.	Security Implementations	Security / access controls implemented, use off firewalls etc.	e.g. Privacy - Encryptions, IBM Security Manager etc.
3.	Scalable Architecture	Scalability of architecture (3 – tier, Micro-services)	Technology used - IaaS, PaaS, SaaS (IBM Cloud).
4.	Availability	Availability of application	Technology used - The Availability of getting used to this software or product design is through by accessing IBM Cognos Analytics and IBM cloud.
5.	Performance	Performance of the application	Technology used - The performance should be fast relaying. This prediction system should be made available in cloud to ensure better accessibility and setting a milestone in providing good quality affordable healthcare.

## 6. PROJECT PLANNING & SCHEDULING

### Sprint Planning & Estimation

<b>Sprint</b>	<b>Functional Requirement (Epic)</b>	<b>User Story Number</b>	<b>User Story /Task</b>	<b>Story Points</b>	<b>Priority</b>	<b>Team Members</b>
Sprint-1	Data Exploration	USN-1	As an analyst,I will explore and analyze the dataset through comparing the paramenters. .	5	High	KOTHURI SAI SUMANTH
Sprint-2	Cleaning and performing EDA's	USN-2	As an Analyst I will developcode for performing EDA's	5	High	KOUSHIGAN L
Sprint-3	Model Classification	USN-3	As an Analyst I can developcode for finding the best fit model	5	High	LOGESHWARAN Y S
Sprint-4	Data Prediction And UI	USN-4	As a Data Analyst, I willcreate code for model prediction and UserINterface	5	High	KIRAN S

## **Sprint Delivery Plan**

<b>Sprint</b>	<b>Total Story Points</b>	<b>Duration</b>	<b>Sprint StartDate</b>	<b>Sprint End Date (Planned)</b>	<b>Story Points Completed(as on Planned End Date)</b>	<b>Sprint Release Date (Actual)</b>
Sprint-1	10	5 Days	24 Oct 2022	29 Oct 2022	10	29 Oct 2022

Sprint-2	10	5 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	10	5 Days	07 Nov 2022	12 Nov 2022	10	12 Nov 2022
Sprint-4	10	5 Days	14 Nov 2022	19 Nov 2022	10	19 Nov 2022

## 7. PROJECT SUMMARY AND RESULT

In the heart attack prediction system, for the research purpose and understanding purpose the values and results obtained from the datasets are depicted as various forms of graphs. Based on the parameters and the clinical output the graphs are generated.

Initially the datasets are led through various EDAs to make it clean and to get deep understanding over the datasets in Jupyter notebook.

Out of 270 datasets, its splited in the ratio of 80:20.Using test\_train\_split 80% of 270 datasets are trained and used as models.Rest of 20% datasets of the csv file are checked whether the target outputs obtained are valid which have been compared and tested with the trained model.

The datasets are trained as logistics regression machine learning model and decision tree classifier model. In logistics regression model, from the trained datasets a curve graph is plotted and a threshold value is marked.

For each testing data sets if the output value is greater than the threshold value the person will not be affected by heart attack or else if the value is below the threshold value for the testing dataset it is concluded that the person will face or affect by heart attack.

After getting to know about datasets and the bestfit algorithm for the project the training and testing phase of model is developed in Flask. Then a interactive FrontEnd is made using Angular to call the endpoints in Flask to display the necessary results.

Also Data Visualization is made over IBM Cognos for better understanding of Datasets. Project Snapshots,

The screenshot displays the user interface of the 'Heart Disease Prediction' application. The background is a solid light purple color. On the left side, there is a main content area with the following elements: a title 'Heart Disease Prediction' in bold black font; a large heading 'Future Is Here, Start Exploring Now.' in bold black font; a paragraph of text stating 'This Heart Disease Prediction System makes user to Visualise and Predict Heart Disease among people. So that users can be conscious about their health. All you need is to know clinical info and details.'; another paragraph stating 'See through current people analysis and current trend of disease before going to prediction'; and a purple button labeled 'PROCEED'. On the right side, there is a white 'Sign In' modal box with a shadow. This box contains two input fields: 'Enter Email Here' and 'Enter Password Here', both with light gray borders. Below these fields is a purple button labeled 'LOG IN'. At the bottom of the modal box, there is a link that says 'New User? Sign Up' in purple text.

## Heart Disease Prediction

### Provide Values

Age

Chest Pain Type

Cholesterol

EKG REsult

Exercise Angia

Slope Of ST

Thallium

Gender

Blood Pressure

Fasting Blood Sugar

Maximum Heart Rate

ST Depression

Number of Vessels Fluro

Predict

## Heart Disease Prediction



**As per your report you are healthy and have less or no chance for facing Heart Disease.**

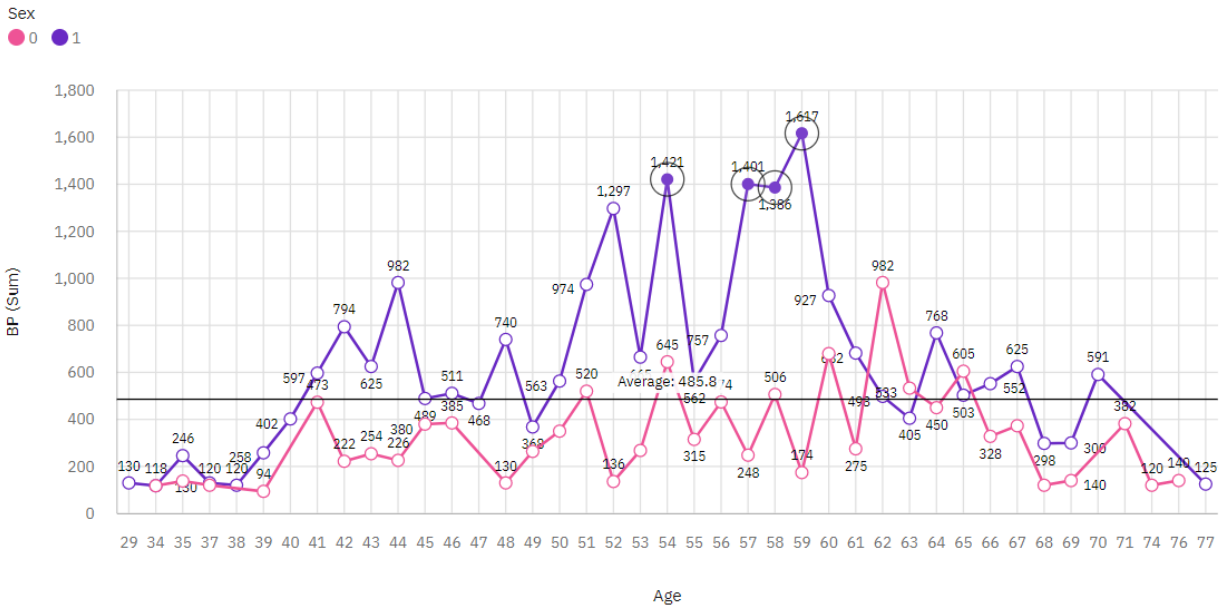
Anyhow, follow the below steps for more healthy life

Don't smoke or use tobacco. One of the best things you can do for your heart is to stop smoking or using smokeless tobacco. ... Get moving: Aim for at least 30 to 60 minutes of activity daily.

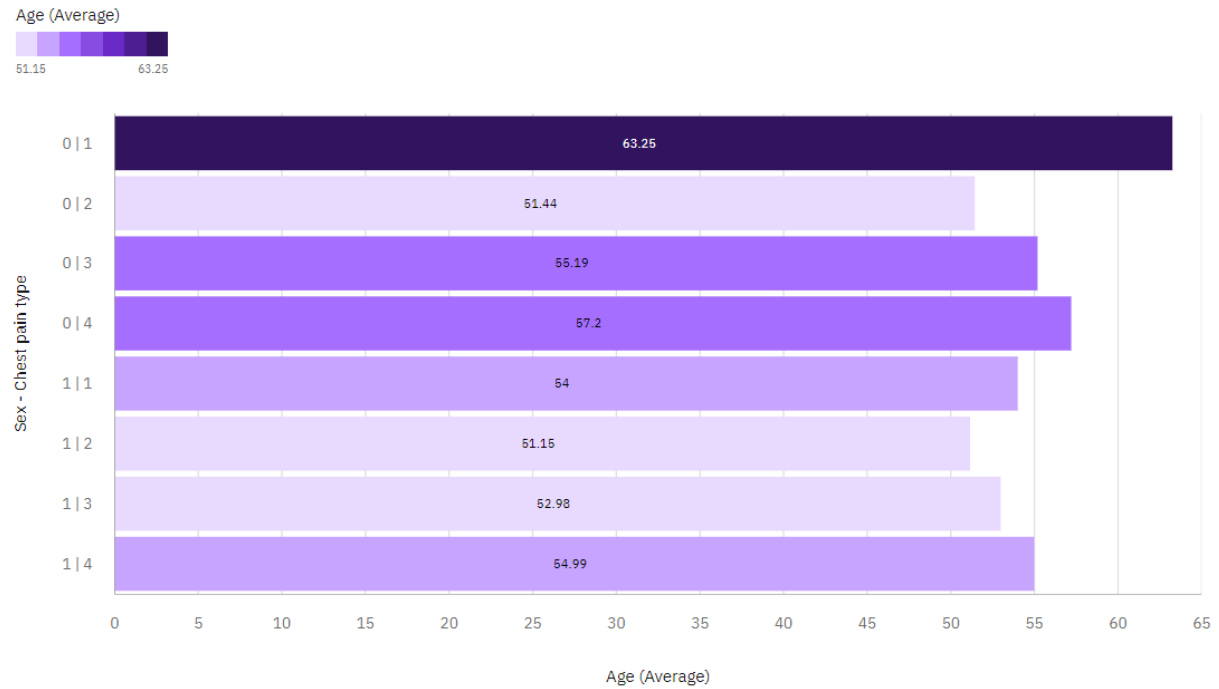
Eat a heart-healthy diet. Maintain a healthy weight. Manage stress. Get regular health screenings.

IBM Dashboard Snapshots,

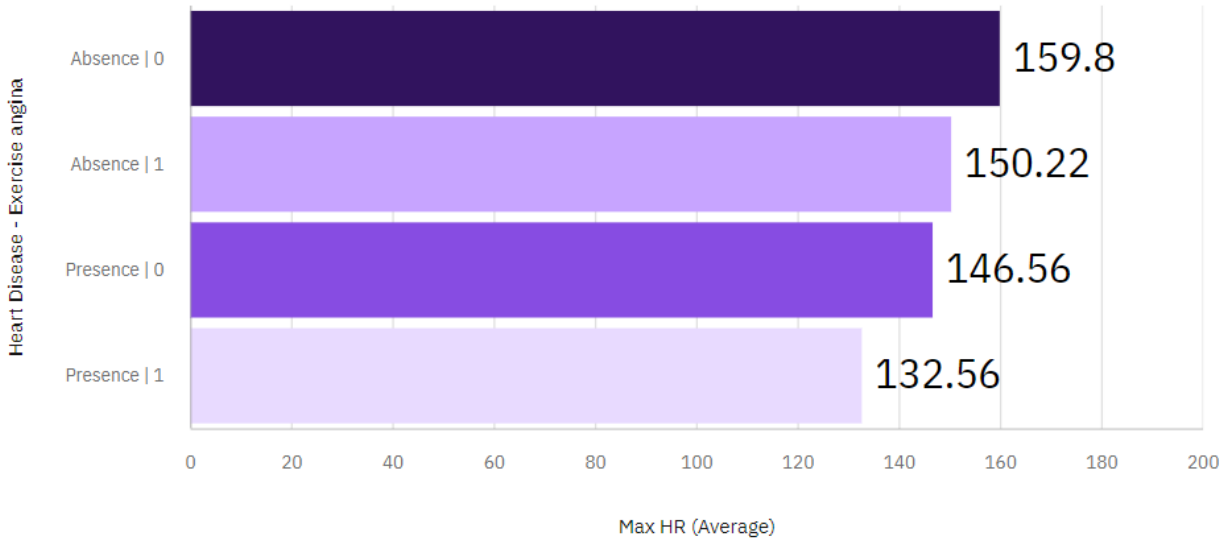
BP by Age colored by Sex



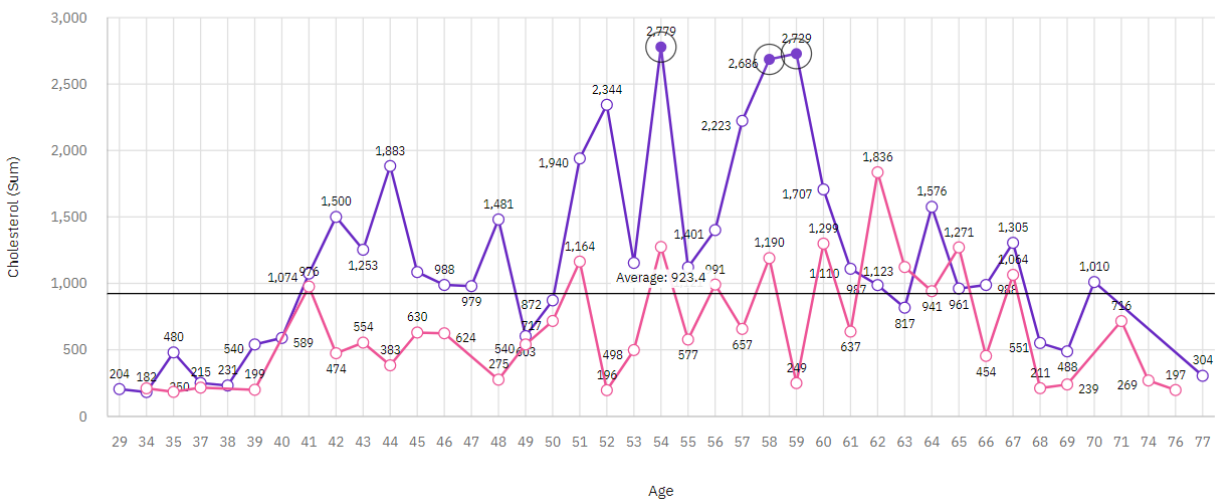
Age by Sex and Chest pain type



Max HR by Heart Disease and Exercise angina colored by Max HR



Cholesterol by Age colored by Sex



## 8. ADVANTAGES & DISADVANTAGES

### ADVANTAGES:

As we are using the pre-existing data of the patient or person, it predicts the chances of heart diseases and classifies the patient based on the risk level. Due to this prediction,

the patients can consult to the respective doctor as a precautionary measure and can be mental free from the thoughts of their problem

The proposed work predicts the chances of Heart Disease and classifies patient's risk level.

User friendly and interactive application.

#### DISADVANTAGES:

Data analytics techniques do not help to provide effective decision making.

Those with heart failure can develop swelling, dizziness, and other symptoms that can affect their ability to complete daily tasks. A person with diagnosed heart disease must also live with the stress of knowing they have a long-term illness that could result in a cardiac event, such as heart attack or stroke. Patients with history of CVD may experience various physical and emotional symptoms such as fatigue, edema, and sleeping difficulties that limit their physical and social activities which will in turn result in poor quality of life.

Cannot handle enormous datasets

Prediction of cardiovascular disease results is not accurate

## 9. CONCLUSION

It is concluded that system will works well and thus it will fulfill the end user's requirement. The system is tested and errors are accurately removed. Heart Disease is one of the leading causes of death worldwide and the early prediction of heart disease is important. The project aims on predicting the heart disease by using the Logistic Regression Algorithm with the help of web application. The probability of disease is found by using certain datasets and the input given by the user.

## 10. FUTURE SCOPE

In this work, we are analyzing the data and predicting the risk of heart disease by merely a history or a data set collected from the person. But, if we want to predict the heart disease from this, there may be risks of being prone to one before the prediction. So, we would like to develop or find some algorithm to analyze data that has been streamingly



collected from the wearable devices like smart watches, fitness bands, healthcare meters and analyze the data in real time to visualize the condition of the heart to the person.

## **11. REFERENCES**

- [1] " Predicting the Risk of Heart Failure With EHR Sequential Data Modelling," Bo Jin, Chao Che.
- [2] "Heart Attack Prediction and Visualization of Contributing Factors Using Machine Learning" by Megha Banerjee, Reetodeep Hazra, Suvranil Saha, Megha Bhushan, Subhankar Bhattacharjee.
- [3] "A novel approach for heart disease prediction using strength scores with significant Predictors" by Armin Yazdani, Kasturi Dewi Varathan, Yin Kia Chiam, Asad Waqar Malik, and Wan Azman Wan Ahmad.
- [4] "Heart Disease Risk Prediction Using Machine Learning Classifiers with Attribute Evaluators" by Karna Vishnu Vardhana Reddy, Irraivan Elamvazuthi, Azrina Abd Aziz, Sivajothi Paramasivam, Hui Na Chua, and S. Pranavanand